DOCUMENT RESUME

ED 201 529 SE 034 936

AUTHOR Gersten, Russell; Carnine, Douglas
TITLE Direct Instruction Mathematics: Longitudinal Evaluation of Over 2,000 Low-Income Children.
PUB DATE Apr 81
NOTE 49p.: Paper presented at the Annual Meeting of the American Educational Research Association (Los Angeles, CA, April 13-17, 1981).

EDRS PRICE MF01/PC02 Plus Postage.
DESCRIPTORS Curriculum Development; Educational Research; Elementary Education; *Elementary School Mathematics; Kindergarten; Learning Theories; *Longitudinal Studies; *Low Income Groups; *Mathematics Curriculum; Mathematics Education; *Mathematics Instruction; *Teaching Methods
IDENTIFIERS Direct Instruction Follow Through Project; *Mathematics Education Research

ABSTRACT

The results from the longitudinal evaluation of over 2000 low income students involved in the Direct Instruction Follow Through Project between kindergarten and third grade are reviewed. Of the ten major education approaches evaluated, students in the Direct Instruction model performed highest in all areas of Math: (1) Computations (66% significant); (2) Problem Solving (55% significant); and (3) Concepts (37% significant). Mean performance for 2897 students was at the fifty-fourth percentile for Total Math, whereas entry scores were at approximately the eighteenth percentile. Data for effectiveness for those entering with low IQ's are reviewed. Finally, relationships between this research and other studies of teacher effectiveness are articulated, and implications for improved curriculum in the elementary grades are presented. (Author)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED201529

Direct Instruction Mathematics: Longitudinal Evaluation of Over 2,000 Low-Income Children

U S DEPARTMENT OF HEALTH, EDUCATION & WELFARE
NATIONAL INSTITUTE OF EDUCATION

THIS DOCUMENT HAS BEEN REPRODUCED EXACTLY AS RECEIVED FROM THE PERSON OR ORGANIZATION ORIGINATING IT. POINTS OF VIEW OR OPINIONS STATED DO NOT NECESSARILY REPRESENT OFFICIAL NATIONAL INSTITUTE OF EDUCATION POSITION OR POLICY

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

RUSSELL GERSTEN

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

Russell Gersten
Douglas Carnine
University of Oregon

Running Head: Direct Instruction

SE 034 936

Abstract

The results from the longitudinal evaluation of over 2000 low income students involved in the Direct Instruction Follow Through project between kindergarten and third grade are reviewed. Of the ten major education approaches evaluated, students in the Direct Instruction model performed highest in all areas of Math - Computations (66% significant), Problem Solving (55% significant), and Concepts (37% significant). Mean performance for 2897 students was at the 54th percentile for Total Math, whereas entry scores were at approximately the 18th percentile. Data for effectiveness for those entering with low IQ's are reviewed. Finally, relationships between this research and other studies of teacher effectiveness are articulated, and implications for improved curriculum in the elementary grades are presented.

Direct Instruction Mathematics: Longitudinal Evaluation of Over 2,000 Low-Income Children

In 1968, the Follow Through Project began an intensive longitudinal evaluation of nineteen different models for teaching low income, primarily minority, students in the primary grades. Follow Through was one of the largest and most expensive social experiments ever conducted (McDaniels, 1975). Over 180 school districts were involved, ranging from rural areas like Flippin, Arkansas, to the Rosebud Sioux Indian reservation in South Dakota to inner city districts in Philadelphia, New York, and Los Angeles.

The group of educational models evaluated was equally diverse, representing a full gamut of philosophies of early childhood education. These included Piagetian-derived approaches such as Weikart's Cognitive Curriculum and the University of Georgia's Mathemagenics Curriculum, an open classroom model, parent-centered approaches emanating from the University of Florida, a psychodynamic approach developed by Bank Street College of Education, as well as behavioral approaches from the Universities of Pittsburgh, Kansas and Oregon. The University of Oregon's Direct Instruction model is the focus of this paper.

Direct Instruction is characterized by a carefully structured educational environment. Instructional materials used by the Model explicitly teach general rules and problem strategies by clearly presenting students with systematic sequences of examples chosen to reduce the likelihood of a child learning a misrule.

Teachers work with students in small groups, monitoring student performance carefully and providing immediate feedback. This methodology is clearly distinct from variants of the "discovery method" used in a variety of commonly used math materials.

Direct Instruction Arithmetic

In recent years the term "direct instruction" has been used in a variety of ways. It, therefore, seems wise at this point to discuss the evolution of the term and its importance as a methodology for teaching mathematics to young children. "Direct Instruction" (with capital letters) refers to the educational model developed by Engelmann, Becker, Carnine, and their colleagues in the late 1960's to teach basic skills, concepts and problem solving strategies to young children. It represents a synthesis of learning theory, behavioral technology and Engelmann's principles of conceptual analysis (Engelmann, 1969; Engelmann & Carnine, in press; Becker & Carnine, 1980). This paper will only focus on the mathematics component of the model.

Direct Instruction assumes that virtually all children can learn mathematics if lessons are designed so that students can readily understand what is being presented and adequate practice with feedback is provided. The eight steps outlined below are needed for constructing an effective instructional program.

1. Specify objectives
2. Devise problem solving strategies
3. Determine preskills
4. Sequence skills
5. Select teaching procedures
6. Design formats
7. Select examples
8. Provide practice and review

In order to efficiently teach rules, strategies and concepts, examples are carefully selected so as to cover the full range of positive examples as well as clearly indicate negative instances. (A more detailed discussion of materials and instructional strategies appears in Direct Instruction Mathematics, by Silbert, Carnine & Stein, 1981.)

In order to master the curricular content, students must receive large quantities of practice with corrective feedback. To achieve this aim, Direct Instruction Follow Through classrooms were arranged so that each student received at least 30 minutes per day in intensive small group instruction in math. Many of the students were taught arithmetic by the classroom paraprofessional aides, who were trained and supervised by sponsor. Lessons were fast paced, eliciting high levels of unison group response. Teachers followed the scripted lessons in the Distar materials.

The Distar instructional materials emphasize the importance of correction procedures. When a child, or group of children, make an error, this is always considered an opportunity for converting this mistake into a positive learning experience. Procedures specified in the teacher's manual indicate how the teacher can remind the child of an appropriate strategy or rule, and if necessary, model the correct answers for the child, and then create situations where the child can successfully solve that problem or similar problems.

The Direct Instruction model also utilizes a mastery learning paradigm--lessons are paced so that all students in each group perform at an 80 to 90 percent accuracy level. Students are given individually administered criterion-referenced tests (by a tester other than the classroom teacher) on material covered during the last two months. Scores on these tests are used to determine whether the child (or group) needs remediation, and whether individual students need to be placed in either a more accelerated or slower paced group.

Thus, Direct Instruction, as used in the Follow Through evaluation and this paper, refers to a comprehensive educational model--involving, at the minimum, at least 30 minutes a day of intensive small group instruction in Distar Arithmetic.

In the late seventies, a second, complementary definition of direct instruction evolved. In reviews of research studies of effective classroom practices, Rosenshine (1976, Note 1) discussed a set of teacher behaviors which appeared to produce greater student achievement.

> "Direct instruction refers to teaching activities focused on academic matters where goals are clear to students; time allocated for instruction is sufficient and content coverage is extensive; student performance is monitored; questions are at a low cognitive level and produce many correct responses; and feedback to students is immediate and academically oriented. In direct instruction, the teacher controls instructional goals, chooses material appropriate for the students' ability level, and paces the instructional episode. Interaction is characterized as structured, but not authoritarian; rather, learning takes place in a convivial academic atmosphere. As we shall see, direct instruction appears to be the most efficient way of achieving skill mastery, just as it is the most efficient way to teach other skills such as ballet, skiing, swimming, or making a fire." (Rosenshine, Note 1, pp. 12-13)

Although a great number of the studies Rosenshine reviewed focused on students from lower socio-economic backgrounds, his conclusions are consistent with those of several other researchers, some of whom investigated other populations (Brophy & Evertson, 1976; Clark, Gage, Marx, Peterson, Stayrock & Winne, 1976; Fisher, Berliner, Filby, Marliave, Cahen & Dishaw, 1980; Soar, 1973; Stallings, 1975).

Rosenshine's definition fits most of the features of Direct Instruction as specified above, although his conception does not really articulate important

issues in curriculum design. The remainder of this paper concentrates only on students from low income areas who were taught Direct Instruction Arithmetic using the principles expounded above, and the Distar curricula materials. Implications for other populations and for other types of related, but different, curricula materials are thus conjectural.

The Nature of Follow Through and Its Evaluation Design

If Follow Through was one of the largest social experiments ever conducted, it also was one of the most controversial (Becker, 1978; Bereiter & Kurland, in press; House, Glass, McLean & Walker, 1978). The U.S. Office of Education contracted with Stanford Research Institute and Abt Associates to evaluate the effectiveness of the ten major sponsors using a longitudinal quasi-experimental design. (A sponsor was an educational organization or university-affiliated program advocating a specific educational model in its sites.)

For each of these sponsors, six to ten representative sites were selected following a sampling design created by the Huron Institute (Haney, 1977). The second and third cohorts of children (those children beginning in 1969 and 1970, the second and third years of the project) were evaluated in depth. The purpose of examining both cohorts II and III was to establish replicability of findings.

Since random assignment of children to treatment was deemed politicallly unfeasible, a quasi-experimental design (Campbell & Stanley, 1966) was utilized; for each Follow Through site chosen, a roughly equivalent local "comparison group" was selected by Stanford Research Institute. (A major flaw, though not in our opinion an unremediable one, was that the local comparison groups tended to be less disadvantaged than the Follow Through children.) Demographic information was collected on all Follow Through and comparison children on the following variables: family

income, ethnicity, mother's education, and primary language spoken at home. In addition, children were pretested on the Wide Range Achievement Test, one of the only academically-based individually administered tests available in 1969.

At the end of the third grade, students were assessed on all subtests of the Elementary form of the Metropolitan Achievement Test (1970). The subtests of most relevance for mathematics education are Math Computation, Math Problem Solving, Math Concepts, and the Total Math composite. Students were also given the subtests in other academic areas, and three affective measures on self concept--the Coopersmith (1967) Self Concept Inventory, a measure of self-esteem, and the "positive" and "negative" scales of the Intellectual Achievement Responsibility Scale (Crandall, Katkowsky & Crandall, 1965), indicating to what extent children attribute positive and negative academic experiences to themselves as opposed to the external environment.

The evaluation addressed two major questions. The first was to assess to what extent Follow Through was superior to conventional educational practices. The second and to many (e.g., Bronfenbrenner, 1975) the most important, was to examine which approach or approaches were most effective in teaching disadvantaged children. As Bereiter and Kurland (in press) indicated, the inclusion of "mismatched" comparison groups probably make the first question unanswerable. But the second should be clearly answerable, and is of great educational import.

In a sense, many of the controversies concerning the "fairness" of the Follow Through evaluation are circumvented when the target is mathematics instruction. Whereas the affective instruments in the battery, and the Language and Reading subtests of the Metropolitan Achievement Test have been criticized as being too narrow in scope (e.g., House et al., 1978;

McLean, Note 2), there is considerably less controversy concerning the content and validity of the Math Problem Solving and Math Computation subtests of the Elementary form of the Metropolitan Achievement Test (Bereiter & Kurland, in press; Wolf, 1978).

This paper will offer an overview of the Follow Through evaluation design conducted by Abt Associates and Stanford Research Institute (SRI) followed the results in mathematics, . both descriptive and statistical. This data will be supplemented by (a) norm-referenced comparisons (Horst, Tallmadge & Wood, 1975) on the Wide Range and Metropolitan Achievement Tests for all children involved in the Direct Instruction Follow Through Model, (b) a study of the later effects of Follow Through, i.e., how 5th and 6th graduates of the program perform in 5th and 6th grades and (c) a study of the relationship between entry IQ and yearly gains in Mathematics achievement.

Method

Subjects

The children in the sample came from twenty low-income school districts representing a broad range of minorities and communities; urban Blacks in East St. Louis, Washington, D.C., and New York City; Native Americans at Rosebud, SD, and Cherokee, NC; children with Hispanic backgrounds in Uvalde and Dimmitt, Texas; rural whites in Flippin, Arkansas, and rural Blacks in Tupelo, Mississippi and Williamsburg County, South Carolina. The children were taught by regular public school teachers and instructional aides who were trained by University of Oregon supervisors in Direct Instruction teaching techniques. In approximately half the communities Follow Through began in kindergarten and continued through third grade; these sites are called Enter-Kindergarten (EK) sites. In the remaining sites, located primarily in the South and the Southwest, Follow Through began in first grade; these are called Entering-First (EF) sites.

The mean entry scores on the Arithmetic subtest of the Wide Range Achievement Test corresponded to the 19th percentile performance for the EK sites (mean standard score of 85.8, SD = 16.1 for a sample of 2472 children). For the EF sites, the mean standard score was roughly the same: 84.0 (with a standard deviation of 14.4) corresponding to the 16th percentile performance for the 2418 children tested. (Becker & Engelmann, 1978).

As previously mentioned, the Abt-SRI evaluation included only a representative sample of 10 (of the 20) sites. However, children in all 20 sites are considered in the supplementary, sponsor-collected analyses, which are summarized later. One other point should be mentioned here. Approximately 20 percent of the sample were non-low income by Follow Through guidelines. Abt Associates included all students in Follow Through, both those considered low income by Follow Through guidelines, as well as the remaining 20 percent. This appears a reasonable procedure considering Abt's use of analysis of covariance procedures. The University of Oregon analyses, which do not rely on covariance adjustments, use only the low income children involved in Follow Through.

Procedures

Children were taught Distar Arithmetic in small groups of six to ten for at least 30 minutes per day by either a teacher or a paraprofessional aide. Additional time was devoted to completion of worksheets. The essential components of the program are detailed in an earlier section of this paper.

The nature of mathematics instruction in the comparison schools is largely unknown. This is rightly cited as a limitation of the evaluation design, though perhaps understandable considering the state of the art in 1968-73. It is highly unlikely, however, that Direct Instruction methods were used in these classrooms.

Measurement

All students in the evaluation sample were tested upon entry into the program on the Wide Range Achievement Test. Children were tested at the end of grades 1, 2, and 3 on the Metropolitan Achievement Test (MAT). The Math Problem Solving and Math Computation sections of the MAT are generally considered reliable (with coefficients over .85) and valid subtests; the validity of the Math Concepts subtests is more open to question (Bereiter & Kurland, in press; Buros, 1978; Wolf, 1978).

The Coopersmith Self-Concept Inventory and IARS were utilized as supplemental measures. The reliabilities of these tests are appreciably lower than the MAT--.69 for Coopersmith; .55 for IARS (House et al., 1978)--though many would argue they are reliable enough to be utilized in comparison of sample means (Nunnaly, 1978; Wisler, Burns & Iwamoto, 1978). In any case, their role is primarily supplemental in this context.

Evaluation Designs

Naturalistic studies and evaluation efforts lend themselves to a variety of design and analyses strategies, partly by choice, partly by necessity. Each of the evaluation designs will be outlined below. They are:

1. Use of local and pooled national comparison groups by the independent evaluators, Stanford Research Institute and Abt Associates for Cohorts II and III.
2. Norm referenced comparisons on all mathematics subtests of the Metropolitan Achievement Test (MAT). Norm-referenced comparisons (Horst, Tallmadge and Wood, 1975) use the standardization sample of an achievement test as the comparison group.

Supplemental analyses include:

1. Relationship between IQ and entry and yearly gains on the MAT.
2. Later effects of Follow Through in grades 5 and 6 (Cohorts II & III).
3. Self-concept measures from the Abt evaluation.

The Abt Longitudinal Analyses

Comparisons with Local and Pooled Controlled Groups. For each site selected for the national longitudinal evaluation sample (Stebbins, St. Pierre, Proper, Anderson, & Cerva, 1977), an analysis of covariance (with multiple covariates) was conducted on each of the math subtests of the MAT, as well as the Total Math composite. The individual child was considered as the unit of analysis. Covariates included SES, mother's education level, race, primary language spoken at home, and entry scores on the Wide Range Achievement Test (WRAT).

An effect was considered significant if:

(1) the F-ratio from the analysis of covariance indicated $p \leq .05$, and

(2) the magnitude of the treatment effect was at least .25 pooled SD units. (This was used to weed out the possibility of finding, at the larger sites, effects that were statistically significant but not large enough to be considered educationally meaningful.) Any effect not meeting these two requirements was considered null, or inconclusive. [1]

Since the authors of the Abt report were well aware of the presence at some sites with somewhat-less-disadvantaged comparison groups, each Follow Through sample was compared to both the local comparison sample and a pooled comparison group, the latter consisting of all the comparison group children in the entire national sample. Both covariance analyses were included in the final results.

Finally, the Abt authors presented what is essentially a "box score" for each of the nine major educational models.[2] For each sponsor, the "box score" was derived in the following fashion:

$$\text{Percentage of Significant Outcomes} = \frac{\text{Number of Positive Effects - Number of Negative Effects}}{\text{Total Number of Possible Effects}}$$

Note that to be considered positive or negative, the effect had to be both statistically significant ($p \leq .05$) and educationally significant. If not, it was considered null. Note that if there were evidence of a clear mismatch between Follow Through and comparison groups that would yield untrustworthy results, the comparison was "grayed out" by Abt, i.e., considered inconclusive. (It is important to indicate here that, though there were many instances of mismatches in the evaluation design, there were also many instances of reasonably equivalent comparison groups.)

Since each sponsor had a different number of sites (e.g., Direct Instruction had 16 sites and thus 32 possible comparisons on each subtest of the MAT, whereas SEDL had 8 possible comparisons), percentages of significant outcomes are presented in Table 1 for each math subtest--Computation, Problem Solving, Concepts, and the Total Math composite. Figure 1 offers a graphic presentation of the results for Total Math.[3]

Insert Table 1 and Figure 1 about here

From Table 1 it is clear that the Direct Instruction model is superior to the other eight approaches evaluated in all domains. The highest incidence of positive effects is in Math Computation, where 66% of the effects are both statistically and educationally significant. On Math Problem Solving, a higher level cognitive skill, the results are nearly as impressive: 55 percent of the effects are positive. The next highest model, SEDL, has only 8% of its effects positive; five of the nine models show negative effects. Thirty-seven (37) percent of the effects are positive in Math Concepts, also a test of higher order arithmetic skills, in which six of the nine models show net negative effects. The results indicate that the Direct Instruction

model produces consistently positive effects in both basic skills and higher order skills in a variety of sites over two cohorts of children. It is the first-ranked of the nine major approaches in all domains tested, including Total Math.

To enhance the reader's understanding of the Abt evaluation, and the range of Follow Through sites, the unadjusted raw scores on each subtest, with percentile equivalents, are presented for cohort III; EK sites are presented in Table 2 (and graphically in Figure 2); EF sites in Table 3.[4] A Plus (+) in the bottom row for each subtest indicates that, when covariance adjustments were made for initial inequities, the effect was assessed to be statistically significant ($p < .05$) and educationally significant (magnitude of the effect was at least 1/4 pooled SD unit).

Insert Table 2, Figure 2 and Table 3 about here

Probably no discussion of the Abt Report is complete without a discussion of variability between sites. The authors of the report (Stebbins, et al., 1977) asserted that this was one of the major findings of the evaluation--that even generally effective sponsors had one or two low-performing sites, and even the least effective sponsors had one or two sites with some positive effects.

It should be clear to the reader by now that, according to Abt's analytic procedures, at least some of the "unexplained" variability is, in fact, explained by variability in the appropriateness or "fairness" of the local comparison groups. For example, for the Tupelo, Mississippi site, where 95 percent of the students were Black with a median family income of $2625, SRI selected a comparison group which was 100 percent white with a median income of $9334.

Not surprisingly, results of the Tupelo comparisons were "not significant." (See Campbell and Erlebacher, 1970, for a detailed discussion of why covariance adjustments cannot adequately deal with severe mismatches.) A similar problem arose in West Iron County, where the median income and entry scores of the comparison group were appreciably higher. On the other hand, the comparison group selected for the Williamsburg County, South Carolina site was fair by most standards; the Follow Through and comparison groups were both over 95 percent black with median incomes within a few hundred dollars of each other and roughly comparable entry scores on the WRAT. Williamsburg County, an EF site in Table 3, gives one a sense of the diversity of communities served. Although the absolute performance of the site is quite low (37th percentile in Computation, 29th in Concepts, and 25th in Problem Solving), the Follow Through children consistently outperform local comparison children, often by as much as one full standard deviation unit. The effects are thus consistently positive, all significant at the .01 level. Williamsburg County was considered the poorest county in the U.S. in the 1970 census with one of the highest levels of illiteracy. Thus, considering the demographic information and the quality of the local education system, the effects are quite impressive.

Another source of variability for the Direct Instruction model, not adequately considered in the Abt report, was that beginning in 1971, one site, Grand Rapids, terminated its relationship with the sponsor two years before the students were tested. Nonetheless, Abt included the Grand Rapids results in the final report. Note that for the EK sites (Table 2 and Figure 2), with the one exception of Grand Rapids, all sites are at or <u>above</u> the national median. Thus, the bulk of the variability is <u>above</u> the national norm level, which is hardly a problem. This type of variability is due to a variety of factors--quality of the teaching

staff, administrative support, community and parental support, amount of local expenditure on education--seemingly a reasonable phenomenon in a diverse society.

Norm referenced comparisons. The Abt report included the mean standard score in Total Math for each sponsor converted to a percentile equivalent. This is reported in Table 4. Each box represents one-fourth of a standard deviation unit. The horizontal line represents mean performance for students

---

Insert Table 4 about here

---

in low income, minority schools in mathematics according to recent USOE evaluations (Ozenne et al., 1974). Note that children in Direct Instruction Follow Through are performing at a level corresponding to the 48th percentile, within two percentile points of the national median and over one standard deviation above the expected level from children in low income minority schools. Note that the Direct Instruction children are performing at a significantly higher level than children in the other Follow Through models, which range from the 11th to 28th percentiles.

Bereiter-Kurland Reanalysis. Though the logic of the Abt analyses appears to be basically sound, the analytic procedures are, to say the least, cumbersome. Bereiter and Kurland (in press) reanalyzed the raw data from the Abt report using the site, not the child, as the unit of analysis, reasoning that the site and not the child was the real locus of the intervention. Because of the inequities in the selection of certain of the local comparison groups, the design could not adequately answer the question of whether, in each community, Follow Through was better than conventional practices. However, the planned variation design does allow one to look at the question of which approaches (sponsors) are significantly more effective in each domain tested. Rather than greyouts and box scores, they used the local comparison group scores on the subtest

of interest as one of the covariates, in an attempt to reduce the bias caused by non-equivalent local comparison groups. A one-way analysis of covariance with multiple covariates was conducted on each subtest between the nine sponsors, using (a) WRAT entry scores, (b) SES, (c) primary home language and (d) local comparison group post-test scores as the four covariates. Where significant differences were found, Neuman-Keuls post hoc tests were performed on the adjusted means. Table 5 presents the results of the Neuman-Keuls tests for the Direct Instruction model on the three math subtests and Total Math. If Total Math, Problem Solving, Concepts, or Computation appears in a box, this indicates a significant ($p < .05$) difference was found between Direct Instruction and that sponsor on the subtest of interest. Note that their findings, using more sophisticated statistical analyses, basically corroborate the findings of the Abt group.

Insert Table 5 about here

Effects of Direct Instruction on Self-Concept

People unfamiliar with direct instruction approaches were surprised by the conclusion that children from behavioral approaches performed better on self-concept measures than children from the approaches whose philosophy was to enhance the self-concept of minority children. The results of the self-concept measures are summarized in Table 6. Direct Instruction is the highest ranked model, followed by the Behavior Analysis model. Note that models such as Bank Street or Open Education (EDC), despite their philosophic orientation, have negative net results.

Insert Table 6 about here

The validity of the affective measures selected in the Follow Through evaluation have been criticized (House et al., 1978; McLean, Note 2). While

it is true that the Coopersmith Self-Esteem Inventory has less than ideal reliability, and there are problems with the formats of both measures, it also seems clear that the reliability of the tests are certainly adequate for research involving large samples (Nunnally, 1978; Wisler et al., 1978) and trends that are replicated with thousands of children in 50 sites over two cohorts of children would appear to be of more than passing interest, and not merely a chance finding.

The implication seems to be that structured, direct instruction approaches that focus on academic goals appear to enhance the self-concept of children. This makes sense. Children who have become competent in math, reading and language might well be expected to feel good about themselves.

The Sponsor-Collected Data

To supplement the independent evaluation, Metropolitan testing was done of all cohorts of children at all sites under the supervision of the Direct Instruction Model. To qualify for the sample a child was required to (a) attend the full four years (three years for EF sites), and (b) be considered low income by Follow Through regulations.

Insert Tables 7, 8 about here

Tables 7 and 8 demonstrate year-by-year scores for low-income students for each Math subtest of the MAT, as well as Total Math. For the EK children (Table 7) one notes consistent yearly growth in Math Computation, from

the 48th percentile at the end of first grade to the 71st percentile at the end of third grade. Math Problem Solving and Concepts are more stable, always within one-fourth of a standard deviation of the national norm. The Total Math score is at the 47th percentile by the end of first grade and increases slightly in third grade. The mean score at the end of the program is slightly above the national norm level. (Note that these scores are within a few percentile points of those independently collected and analyzed by Abt in Table 4.)

A similar pattern emerges for the Entering-First Follow Through children, though the scores tend to be approximately one-fourth of a standard deviation lower for these children, who experience one less year of instruction.

It seems clear that in the area of mathematics, the model has reached its objective of having low income students perform at grade level. Stebbins et al. (1977) summarized the results for children who had participated in The Direct Instruction approach for four years:

> The Direct Instruction Model is specific in stating that children participating in the Follow Through program are expected to, on the average, perform at the same level as their middle-class peers by the end of third grade. This goal has largely been achieved.. (Vol. IV, pp. A-168 and A-169).

Some data are available on non-low income children who are not from low-income backgrounds. These children attended Follow Through classrooms and were tested along with the low-income children. The non-low-income children had also participated in the program for four years and were tested at the end of third grade.[5] On the Metropolitan computation subtest the mean performance for the 321 students was 4.76, almost a full year ahead of the norm, corresponding to 83rd percentile performance. The mean for the concepts subtest was 4.42 (corresponding to the 68th percentile), and 4.26 for problem-solving (75th percentile).

The mean grade level for Total Math was 4.43, corresponding to the 71st percentile. Although the absence of data for non-low income comparison groups precludes any definitive conclusions, the results certainly suggest that direct instruction can be effective with a broad range of children, not just those who are identified as slow learners.

The Low IQ Study

Direct Instruction hypothesizes that, with a properly implemented mastery learning model, all children, regardless of entry skill level should make roughly yearly gains. To explore this hypothesis, the yearly Total Math scores on the MAT were reanalyzed to explore whether there was a significant interaction between the child's IQ at entry (measured by the Slosson Intelligence Test, an abbreviated form of the Stanford-Binet) and yearly growth on the Total Math subtest of the MAT. For the purposes of the analysis, children were segregated into one of six IQ blocks: below 70, 70-90, 91-100, 101-110, 111-130, over 130. A 6x3 mixed analysis of variance was performed with one between group factor (IQ block) and one within group factor (year of test). For the EK sample, no significant interaction was found, despite the large sample size of 2100 and consequent high power. Those entering with low IQ scores tended to have lower entry scores on the WRAT and lower MAT scores at the end of first grade, but they still maintained the same average growth rate of 1.0 Grade Equivalent units per year. (All analyses were performed using standard scores, and only later converted to percentile ranks and grade equivalents.)

A significant interaction was found for the EF group during third grade. Scheffé-post-hoc tests indicated the low IQ sample (those with IQ's below 70) gained significantly less than those with IQ's above 110. No other significant differences were found. One interpretation of this

finding is that, particularly for those entering with very low skill levels, the year of kindergarten instruction is particularly essential. Other interpretations are certainly possible. In any case, these analyses indicate that with this model, all children--not just the upper two-thirds--can progress at a reasonable rate without the need for special education/remediation services.

Later Effects of Direct Instruction

In 1975 and 1976, an attempt was made to ascertain to what extent effects of the program were enduring. Fifth and sixth grades in seven sites (five in 1976) were tested on all subtests of the Intermediate Level of the MAT. Scores were contrasted with the performance of local comparison children, using analysis of covariance techniques with SES, parent education, race, and primary home language as covariates. Analyses were performed on Math Problem Solving, Concepts, Computation and Total Math. Due to the small sample sizes, results were classified as (a) significant ($p < .05$), (b) indicative of a possible positive effect ($.05 < p < .15$) and (c) not significant. Details of the study are reported in Becker and Engelmann (1978) and Becker and Gersten (Note 4). The results seemed to indicate consistent positive findings in the area of Math Problem Solving, but not in Computation or Concepts. It appears that students retain some of the general problem solving strategies taught in Distar Arithmetic programs, and use them in the fifth and sixth grade material they encounter. However, they are not learning the new computational skills in multiplication, division and complex addition and subtraction, which are supposedly taught after the students leave Follow Through.

Discussion

One possible summary statement about the data might be: "Low socio-economic primary-grade students who received direct instruction tended to perform better in math than students who received other approaches ." A second, probably more important summary statement is that Direct Instruction students achieved at a level much higher than one would predict based on their demographic characteristics and performance upon entering school. The Direct Instruction students in the sponsor data summary were all poor and almost minority students, yet they performed about as well as the test standardization sample of the 1970 MAT, who were primarily white and middle class. These two summary statements suggest two broad questions: "What are the specific components of the Direct Instruction Model that were responsible for the results?" and "How generalizable are the Direct Instruction procedures?"

A list of the possible components of the model responsible for the success can be discussed under four relatively broad categories:

1. The amount of time students spent engaged in mathematics (also known as academic engaged time (Rosenshine, 1976) or academic learning time (Fisher et al., 1980) or opportunity (Cooley & Leinhardt, 1980).
2. The teacher performance variables that are part of the model: systematic correction of student errors, high student accuracy rate, use of reinforcement, rapid pacing.
3. The instructional design procedures embedded in the Distar Arithmetic curriculum.
4. The continual monitoring of student progress via biweekly administration of criterion-referenced tests , as well as the monitoring of teacher performance by structured weekly observations, inservice and preservice workshops (See Morimitsu, Note 3, for details.)

The academic engaged time component is a recurrent issue in recent teacher effectiveness studies. Consistently, the amount of time students are observed to be actively engaged in academic activities correlates with student academic gain both in compensatory education settings (Cooley & Leinhardt, 1980; Rosenshine, 1976; Stallings, 1975) and regular education settings (Fisher, et al., 1980).

Research on the second component has demonstrated that the proper use of various Direct Instruction teaching techniques enhances learner performance: appropriate correction of mistakes (Carnine, in press a; Siegel, 1977), rapid pacing of tasks (Carnine, 1976) and cues to encourage learner participation (Carnine, in press b). Moreover, research has indicated that training in corrections (Siegel, 1977) and in cueing and pacing (Carnine & Fink, 1978) are necessary if the techniques are to be used properly. For example, Siegel (1977) reported that, after training, teachers who were initially "low correction" implementers (i.e., rarely used correction procedures) actually surpassed their higher performing peers. Moreover, the students of the low correction implementers who received training outperformed students of the teachers who were not trained. A naturalistic study conducted in an urban Follow Through setting indicated that classrooms in which teachers showed high rates of (a) correction of student errors, (b) pacing of lessons, (c) mean student accuracy during lesson, (d) use of formats and signals gained more on standardized achievement measures (Gersten, Carnine & Williams, in press) than those teachers with low rates. These findings suggest that the end of third grade Follow Through scores resulted in part from teacher training on direct instruction teaching techniques.

Research has also been conducted on the third component: principles of curriculum design. These principles are embodied in instructional procedures for teaching discriminations, rules, and complex cognitive operations; selecting and sequencing examples; identifying and teaching preskills; using prompts and fades; etc. The basic goal of maximizing learner success is realized by (1) making explicit as early as possible in instruction relevant concept features and step-by-step, general case strategies for working a wide range of complex problems, and (2) beginning instruction in a highly structured context and then systematically moving the learner to unstructured applications. The importance of teaching explicit general case strategies is illustrated in two recent studies; one on teaching a multiplication algorithm (Carnine, 1980) and one on teaching fact relations (Carnine & Stein, in press). Additional research and an elaboration of the instructional design procedures are found in Theory of Instruction (Engelmann & Carnine, in press) and in "Learning and Cognition" in the Thirty-seventh Yearbook of the NCTM (Stevenson, 1975).

Although the instructional design principles manifested in the Distar program are undoubtedly major contributors to student achievement, the relative importance of the Distar material in comparison to teacher implementation, the use of quality control procedures, and amount of academic engaged time has not been investigated. Research on the relative importance of these variables is difficult, but necessary.

An interesting perspective on the situation emerges if one reexamines the data in Table 1. It is clear that both the Behavior Analysis and the Direct Instruction models are effective in teaching skills in Math compu-

tation with 55 percent and 66 percent of significant outcomes, respectively. In the higher-order conceptual operations, the Direct Instruction model also produces significant gains: +55 percent for Problem Solving and +37 percent for Concepts. Behavior Analysis, on the other hand, produces negative values of -13 and -5 percent. Both Behavior Analysis and Direct Instruction utilize principles of extensive time on-task (Stallings, 1975), reinforcement, and consistent monitoring of student performance; yet the Behavior Analysis model utilizes the math curriculum currently used in the school district, whereas Direct Instruction classrooms use the Distar Arithmetic series, which is based on the explicit teaching of general case strategies. There is thus some evidence suggesting that, when amount of engaged time is kept constant, learning higher order cognitive skills is enhanced by a curriculum that explicitly tries to do so.

The discussion thus far was intended to point out several variables that contributed to the students' math achievement. Consideration of each of these sets of variables is critical. The Distar programs are merely procedures for teaching mathematics. Without proper teacher implementation, some form of quality control monitoring, and adequate instruction time, the results reported in this paper could probably not be replicated. Although positive effects may be achieved with lower levels of teacher implementation, more limited monitoring, and less time, drastic reductions in any of these areas would probably adversely effect learner outcomes.

Another major question is "Do Direct Instruction procedures work for children other than those in Follow Through?" Direct Instruction procedures have proven relatively successful with lower performing, primary grade students from various school districts across the United States when

outcomes were measured on standardized achievement tests. High levels of achievement were reported in almost all the eighteen cooperating school districts from across the United States, which included rural and urban settings and children of various ethnic make-ups--American Indian, Black, Mexican-American, Spanish-American, and white (Becker, 1978). The effects of direct instruction on average and above average primary grade students require further research. Guthrie (1977) reported that the Direct Instruction model was the only model to have consistent effects for both low and middle income children enrolled in Follow Through. Although the data on non-low income Follow Through students suggested positive outcomes, conclusions must be withheld since control groups were not included. However, a study by Adamson (1976) with a control group, illustrates the identical power of Distar's problem solving approach with above average students. (The students had an average IQ of 111.) The Distar program was compared with Developing Mathematical Processes (DMP) which "relies heavily on perceptual activities, discovery, and open-ended questions with a manipulative approach utilizing measurement" (p. 70). The findings were that

> "Many of the children in both groups appeared perplexed and asked for help when this section (the Stanford Achievement Test) was introduced. However, they were all simply told to try and to do the best they could. The DISTAR group averaged 6.52 problems correct, while the DMP group averaged 3.00 problems correct, which appears to imply that the DISTAR group was able to transfer from a known form of computation to an unknown form, while the children in the DMP group were unable to do so. This may be understandable when one considers that the primary emphasis in DISTAR, but not in DMP, is on computation. However, looking at the attempted scores, a discrepancy seems to remain in the fact that problem-solving has such emphasis in the DMP program, yet so many children (with an average four points higher IQ) were hesitant to attempt a new situation" (Adamson, p. 72).

An interesting side finding was that Distar worked equally well with cooperative and non-cooperative children, but DMP did not. "... [i]n the DMP program, the uncooperative child can choose not to learn, while in the Distar program the child is not given the choice not to learn, and therefore he (sic) learns." (p. 75).

Besides generalization across a variety of student backgrounds and income levels, generalization over time is also important. There is evidence that the effects of the model in the area of math problem-solving did persist into the intermediate grades. The effects of the new Direct Instruction programs designed for fourth, fifth and sixth graders will require further research and will hopefully reflect achievement levels higher than those reported for students who had gone into non-direct instruction classrooms after third grade.

At the present time there are several limitations in generalizing the present results to other less advantaged learners, grades other than the primary grades, and activities other than those covered in standardized achievement tests. Measures need to be devised and validated before research on math applications other than those found in standardized achievement tests can be carried out. However, the data do document the capability of low income children to succeed in mathematics, when receiving direct instruction. As the variables or factors that may account for these instructional effects are systematically investigated with other populations, grades, and activities, the generalizability of the present finds may be extended quite dramatically.

Reference Notes

1. Rosenshine, B. Direct instruction for skill mastery. Paper presented at the University of Wisconsin at Milwaukee, Wisconsin, April, 1979.
2. McLean, L. D. Evaluation of early childhood education--No simple answer to the right questions. Unpublished manuscript. Toronto, Ontario: Ontario Institute for Studies in Education, 1979.
3. Morimitsu, C. Supervision of direct instruction. Eugene, OR: University of Oregon, 1979.
4. Becker, W. C. & Gersten, R. A follow-up of Follow Through. Paper submitted to Society for Research on Child Development, Eugene, OR, 1980.

References

Adamson, G. Mathematics achievement between first-grade students using developing mathematical processes and DISTAR arithmetic mathematics instruction (Doctoral dissertation, Brigham Young University, 1975). Dissertation Abstracts International, 1976, 36, 4211A. (University Microfilms. No. 76-683)

Becker, W. C. The national evaluation of Follow Through--Behavior-theory-based programs come out on top. Education and Urban Society, 1978, 10, 431-458.

Becker, W. C. & Carnine, D. Direct Instruction: An effective approach to educational intervention with the disadvantaged and low performers. In B. Lahey & A. Kazdin (Eds.), Advances in Clinical Child Psychology. New York: Penguin Press, 1980.

Becker, W. C. & Engelmann, S. Analysis of achievement data on six cohorts of low-income children from 20 school districts in the University of Oregon Direct Instruction Follow Through Model. Technical Report #78-1. University of Oregon, Follow Through Project, Eugene, OR, 1978.

Becker, W. C., Engelmann, S. E., Carnine, D. & Maggs, A. Direct instruction technology--making learning happen. In P. Karoly and J. J. Steffen (Eds.), Advances in Child Behavior Analysis and Therapy: Vol 2. New York: Gardner Press, 1980.

Bereiter, C. & Kurland, M. Were some Follow Through models more effective than others? Interchange. in press.

Bronfenbrenner, U. Is early intervention effective? In B. Friedlander, C. Sterritt, & G. Kirk (Eds.), Exceptional Infant Assessment and Intervention (Vol. 3). New York: Bruner/Mazel, 1975.

Brophy, T. & Evertson, C. M. Learning from teaching: A developmental perspective. Boston: Allyn & Bacon, 1976.

Buros, O. Eighth mental measurement yearbook. Highland Park: Crypton Press, 1978.

Campbell, D. T. & Erlebacher, A. How regression artifacts in quasi-experimental evaluations in compensatory education tend to underestimate effects. In J. Hellmuth (Ed.) Disadvantaged child: Compensatory education: A national debate (Vol. 3). New York: Brunner/Mazel, 1970.

Campbell, D. T. & Stanley, J. Experimental and quasi-experimental designs for research. Chicago: Rand McNally, 1966.

Carnine, D. Effects of two teacher presentation rates on off-task behavior, answering correctly, and participation. Journal of Applied Behavior Analysis, 1976, 9, 199-206.

Carnine, D. W. Preteaching versus concurrent teaching of the component skills of a multiplication algorithm. Journal of Research in Mathematics Education, 1980, 11, 375-378.

Carnine, D. Correcting word identification errors of beginning readers. Education and Treatment of Children, in press (a).

Carnine, D. High and low implementation of direct instruction teaching techniques. Education and Treatment of Children, in press (b).

Carnine, D. & Stein, M. Organizational strategies and practice procedures for teaching basic facts. Journal of Research in Mathematics Education, in press.

Carnine, D. & Fink, W. T. Increasing the role of question-asking and use of signals in direct instruction trainees. Journal of Applied Behavior Analysis, 1978, 11, 35-66.

Clark, C. M., Gage, N. L., Marx, R. W., Peterson, P. L., Stayrock, N. C., & Winne, P. H. A factorially designed experiment on teacher structuring, soliciting and reading. Final report. Stanford, CA.: Center for Research and Development in Teaching, 1976.

Cooley, W. W. & Leinhardt, G. The instructional dimensions study. Education Evaluation and Policy Analysis, 1980, 2, 7-25.

Coopersmith, S. The antecedents of self-esteem. San Francisco: W. H. Freeman, 1967.

Crandall, V. C., Katkowsky, W., & Crandall, V. J. Children's beliefs in their own control of reinforcements in intellectual-academic achievement situations. Child Development, 1965, 36, 91-109.

Elashoff, J. D. Analysis of covariance: A delicate instrument. American Educational Research Journal, 1969, 6, 383-401.

Engelmann, S. Conceptual learning. San Rafael, California: Dimensions Publishing, 1969.

Engelmann, S. & Carnine, D. A direct instruction perspective. Educational Technology, in press.

Fisher, C. W., Berliner, D. C., Filby, N. N., Marliave, R., Cahen, L. S., & Dishaw, M. M. Teaching behaviors, academic learning time and student achievement: An overview. In C. Denham & A. Lieberman (Eds.), Time to learn. Washington, D. C.: USOE/ NIE Printing, 1980.

Gersten, R., Carnine, D. & Williams, P. Measuring implementation of a structural educational approach: An observational approach. Educational Evaluation and Policy Analysis, in press.

Guthrie, J. T. Research views - Follow Through: A compensatory education experiment. The Reading Teacher, 1977, 3, 240-244.

Haney, W. A technical history of the national Follow Through evaluation. Sponsored by OE/DHEW, Cambridge, Mass.: Huron Institute, August, 1977.

Horst, D. P., Tallmadge, G. K., & Wood, C. T. A practical guide to measuring project impact on student achievement. Monograph No. 1 on Evaluation in Education. Washington, D.C.: U.S. Government Printing Office, 1975.

House, E. R., Glass, G. V., McLean, L. D., & Walker, D. F. No simple answer: Critique of the "Follow Through" evaluation. Harvard Educational Review, 1978, 28, 128-160.

McDaniels, G. L. The evaluation of Follow Through. Educational Researcher, 1975, 4, 7-11.

Metropolitan Achievement Tests. New York: Harcourt, Brace, Jovanovich, 1970.

Nunnally, J. Psychometric theory (2nd ed). New York: McGraw Hill, 1978.

Ozenne, D. et al. United States Office of Education. Annual Evaluation Report on Programs Administered by the U.S. Office of Education, FY 1975. Washington, D.C.: Capital Publications, Educational Resources Division, 1974.

Rosenshine, B. Classroom instruction. In N. L. Gage (Ed.), Psychology of teaching: The 77th yearbook of the National Society for the Study of Education. Chicago: National Society or the Study of Education, 1976.

Siegal, M. A. Teacher behavior and curriculum packages: Implications for research and teacher education. In L. J. Rubin (Ed.), Handbook of Curriculum. New York: Allyn & Bacon, 1977.

Silbert, J., Carnine, D., & Stein, M. Direct instruction mathematics. Chicago: Charles Merrill, 1981.

Soar, R. S. & Soar, M. R. An empirical analysis of selected Follow Through programs. An example of a process approach to evaluation. In I. J. Gordon (Ed.), Early Childhood Education. Chicago: University of Chicago Press, 1973.

Stallings, J. A. Implementation and child effects of teaching practices in Follow Through classrooms. Monographs of the Society for Research in Child Development, 1975, 40.

Stebbins, L. B., St. Pierre, R. G., Proper, E. C., Anderson, R. B., & Cerva, T. R. Education as experimentation: A planned variation model. Volumes IV-A-D. An evaluation of Follow Through. Cambridge, Mass.: Abt Associates. 1977.

Stevenson, H. W. Learning and cognition. In J. N. Payne (Ed.), Mathematics learning in early childhood. Reston, Virginia: National Council of Teachers in Mathematics, 1975.

Wisler, C., Burns, J. & Iwamoto, D. Follow Through A response to the critique by House, Glass, McLean & Walker. Harvard Education Review, 1978, 48, 171-185.

Wolf, R. Review of Metropolitan Achievement Test. In O. K. Buros (Ed.) The sixth mental measurements yearbook. Highland Park, NJ: The Gryphon Press, 1978.

Footnotes

1. In the four year interim between the design of the study by SRI (1967 and 1968) and the analysis of the data by Abt Associates, a growing awareness among statisticians of the dangers of using the covariance adjustments advocated by Campbell and Stanley (1963) emerged. (Campbell & Erlebacher, 1970; Elashoff, 1969). Campbell and Ehlrbacher (1970) demonstrated how, in quasi-experimental designs such as Follow Through or Head Start evaluations, covariance adjustments tend to systematically underadjust for the more deprived group (i.e., usually the Follow Through or Head Start children). Thus, due to the underadjustments, treatment effects that are truly positive may yield non-significant F-ratios; and cases where the treatment is ineffectual (though not harmful) may yield F-ratios indicating significant, negative effects. To safeguard against these faulty inferences, the authors of the Abt Report (Stebbins et al., 1977) grayed out - that is, considered untrustworthy - any cases when the Follow Through and comparison group differed by more than one-half of a standard deviation unit on a demographic covariate.

2. When considering the results, a few reminders might facilitate understanding of this highly complex design. Nine of the 20 Direct Instruction sites were selected by the independent evaluation agencies to participate in the intensive longitudinal evaluation of cohort II. Cohort II children began in 1969. In some sites, Follow Through began in kindergarten; in others, it began in first grade. Children in the former, referred to as EK sites, completed Follow Through in the spring of 1973 when they were tested on the Metropolitan Achievement Test. Children in EF (Entering-First) sites finished grade 3 in 1972. Seven of the nine sites were selected for a replication study; Cohort II

students began Follow Through in the fall of 1970. To further complicate matters, each local Follow Through sample was compared (via analysis of covariance) with the post-test scores of both the local comparison group and a pooled national comparison group. The major reason the pooled sample was used was to offer a fairer standard for those sites with "mismatched" local comparison groups.

3. The present analysis differes from Abt's in that the "gray out" rules have been revised in the following fashion. If there is a discrepancy of over .5 SD units between Follow Through and comparison group on a covariate favoring the comparison group, and the analysis of covariance indicates a positive effect for Follow Through, the result is not grayed out. Campbell and Erlebacher clearly indicate the covariance adjustment should handicap the Follow Through sample; if the Follow Through children--with this handicap--still significantly outperform the comparison children, there is clear evidence of a strong effect. (See Becker, 1978, for further discussion.)

4. A roughly comparable picture emerged for cohort II.

5. The mean grade equivalents (and percentiles conversions) represent slight overestimates.

Table 1. Percentage of Statistically and Educationally Significant Outcomes Favoring Follow Through Sample
(Adapted from Stebbins et al., 1977, Education as Experimentation, Vol. IVA)

		Metropolitan Achievement Test			
Approach	Sponsor	Math Computation	Math Problem Solving	Math Concepts	TOTAL M
Direct Instruction	University of Oregon	+66%	+55	+37	+50
Behavior Analysis	University of Kansas	+45	-13	- 5	+ 9
Parent Education	University of Florida	+ 5	0	-10	0
Bilingual	Southwest Educational Development Lab (SEDL)	+17	+8	+27	+21
Child Centered, Responsive Education	Far West Lab	+8	0	0	0
Child Development, Psychodynamic	Bank Street	-11	-11	-13	-22
Language Experience (TEEM)	University of Arizona	+8	-17	-26	-26
Piagetian	High Scope Foundation	-7	-17	-17	-2
Open Education	Educational Development Center (EDC)	-24	-14	-20	-25

Table 2. Unadjusted Raw Scores for Follow Through and Local Comparison Group Children on Metropolitan Achievement Test - Cohort III, Entering-Kindergarten (EK) Sites[a]

	NY, NY		Flint, MI		Providence, RI		West Iron Co. MI		Grand Rapids, MI[b]	
	FT	Comp	FT	Comp	FT	Comp	FT	Comp	FT	Comp
MATH COMPUTATION										
Mean Raw Score ($\bar{X}$)	29.48	19.35	27.86	21.17	26.69	18.88	25.53	26.81	16.42	17.36
SD	6.46	7.85	7.60	6.85	7.45	8.48	5.11	5.81	6.12	6.65
Percentile (%ile) Equivalent	80th	35th	73rd	43rd	69th	37th				
Was effect significant?[c]	+(Yes)		+		+		64th	69th	24th	28th
MATH CONCEPTS										
$\bar{X}$	22.32	17.52	21.64	18.07	20.12	18.41	19.88	27.98	15.55	17.86
SD	6.53	7.21	7.77	6.06	7.46	9.13	5.11	5.78	6.95	7.61
Percentile (%ile) Equivalent	52nd	32nd	48th	34th	45th	36th	42nd	76th	26th	33rd
Was effect significant?[c]	+				+					
MATH PROBLEM SOLVING										
$\bar{X}$	18.45	14.25	18.43	13.49	16.18	17.38	15.00	22.26	11.99	12.76
SD	4.93	6.70	7.17	5.81	7.52	7.40	6.26	7.38	5.32	6.01
%ile Equivalent	50th	32nd	50th	28th	41st	46th	36th	63rd	22nd	25th
Was effect significant?[c]	+		+		+					
TOTAL MATH[d]										
Median %ile	66th	32nd	62nd	32nd	52nd	32nd	48th	77th	18th	20th
Grade Equivalent	4.2	3.2	4.1	3.3	3.8	3.3	3.7	4.7	3.0	3.0
Percent of Students 1 or more Years Below Grade Level	6.5	28	9.5	24.4	13.7	36.7	0	2.3	40.7	34.9
Was effect significant?[c]	+		+		+					

a. Adapted from Stebbins et al., 1977, Vol. IVC, pp. 261-276.

b. Direct Instruction model not implemented at this site.

c. Educationally and statistically significant

d. Note that the convention used by Abt changes for Total Math

Table 3. Unadjusted Raw Scores for Cohort III Follow Through and Comparison Groups, Entering-First (EF) Sites[a]

	Dayton, OH		Tupelo, MS		Williamsburg Co., SC	
	FT	Comp	FT	Comp	FT	Comp
MATH COMPUTATION						
Mean Raw Score ($\bar{X}$)	24.53	21.98	24.03	27.68	19.71	12.56
SD	7.88	8.39	8.79	5.92	6.92	5.59
%ile Equivalent	60th	48th	58th	73rd	37th	10th
Was effect significant?[b]	+				+	
MATH CONCEPTS						
$\bar{X}$	19.78	18.05	19.25	26.16	16.33	9.78
SD	7.51	7.57	7.06	6.32	6.28	3.19
%ile	41st	34th	39th	72nd	29th	7th
Significance[b]	+				+	
MATH PROBLEM SOLVING						
$\bar{X}$	15.77	14.53	16.80	21.78	12.77	10.0
SD	6.42	5.80	6.92	6.93	5.7	3.6
%ile	44th	38th	46th	61st	25th	14th
Significance[b]	+				+	
TOTAL MATH[c]						
Median %ile	44th	40th	48th	70th	28th	6th
Grade Equivalent	3.6	3.5	3.7	4.4	3.2	2.3
% of students 1 or more years below grade level	13	27.1	19.7	4.1	32.9	77.8
Significance[b]	+				+	

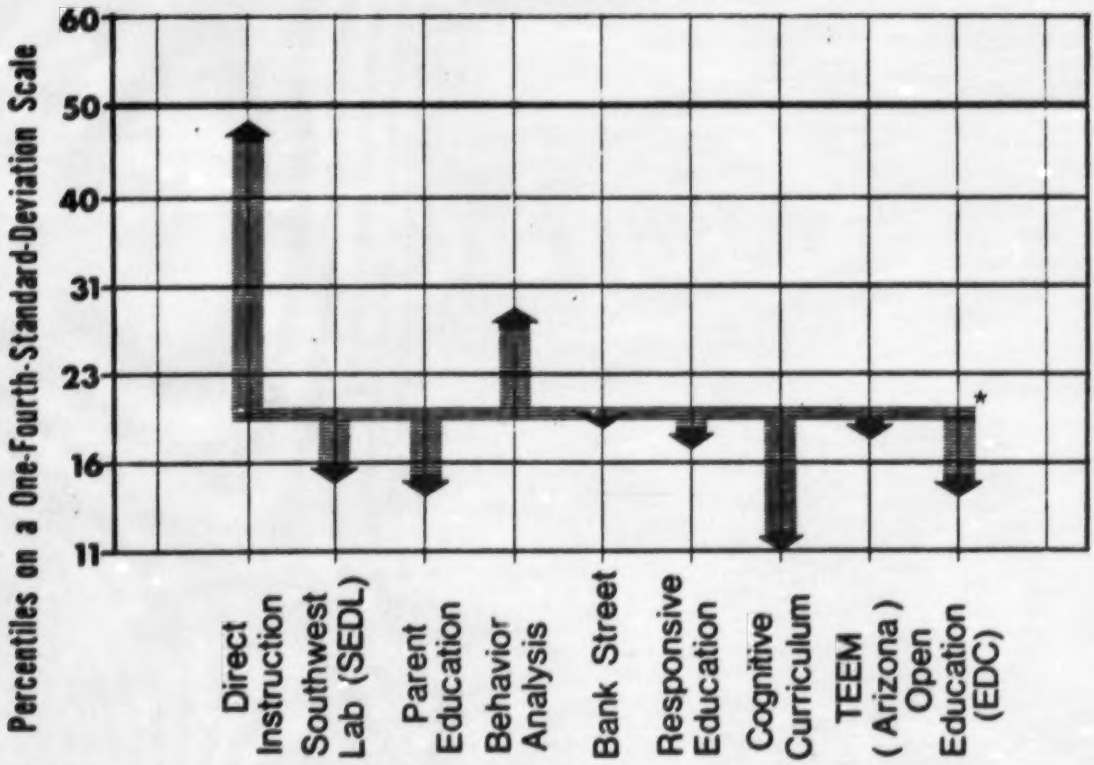
a. Adapted from Stebbins et al., 1977, Vol. IVC, pp. 308-332.

b. + means effect was significant at the .05 level and the magnitude of the effect was at least one-fourth pooled SD unit

c. Note the convention changes for Total Math

Table 4

M.A.T. Total Math



	Direct Instruction	Southwest Lab (SEDL)	Parent Education	Behavior Analysis	Bank Street	Responsive Education	Cognitive Curriculum	TEEM (Arizona)	Open Education (EDC)
%-tile	48	15	14	28	19	17	11	18	14
S. S.	70.1	58.5	57.8	63.8	60.6	59.3	55.8	60.0	58.1
G. E.	3.71	2.85	2.78	3.24	3.03	2.93	2.58	3.00	2.81

Grade equivalent for 50th percentile is 3.75

* USOE reported average performance for low income minority schools (Ozenne et al., 1974)

Table 5

Subtests on which there are Significant Differences Between Direct Instruction and Other Follow Through Sponsors[a]

Parent Education			Math Comp	Total Math
Behavior Analysis				
Bank Street			Math Comp	Total Math
Responsive Education	Problem Solving	Concepts	Math Comp	Total Math
Cognitive Curriculum			Math Comp	Total Math
TEEM (Arizona)		Concepts	Math Comp	Total Math
Open Education (EDC)	Problem Solving	Concepts	Math Comp	Total Math

[a] $p < .05$ by analysis of covariance.
Adapted from Bereiter & Kurland, in press.

Table 6. Percent (and Direction) of Significant Outcomes for Affective Measures (Coopersmith, Intellectual Achievement Responsibility Scale)

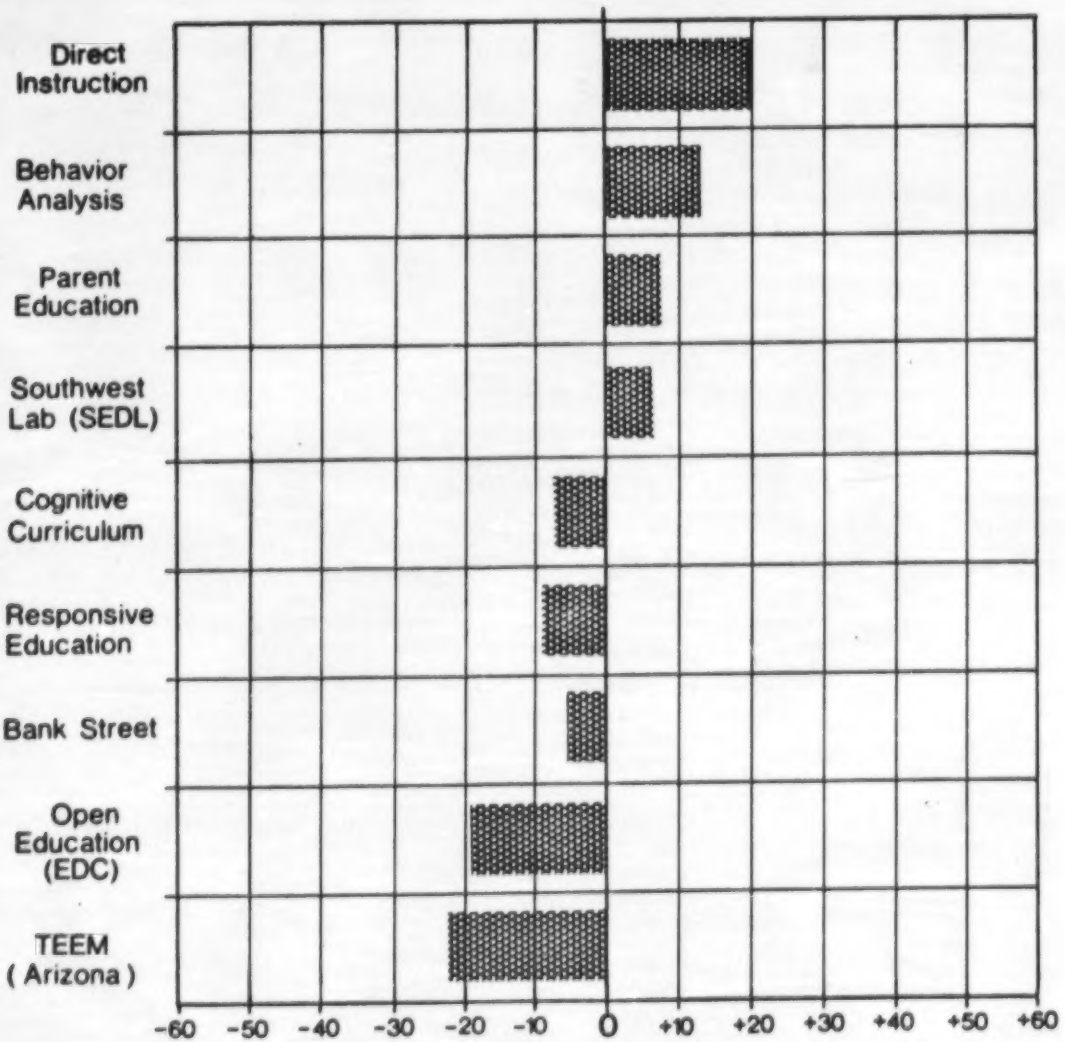
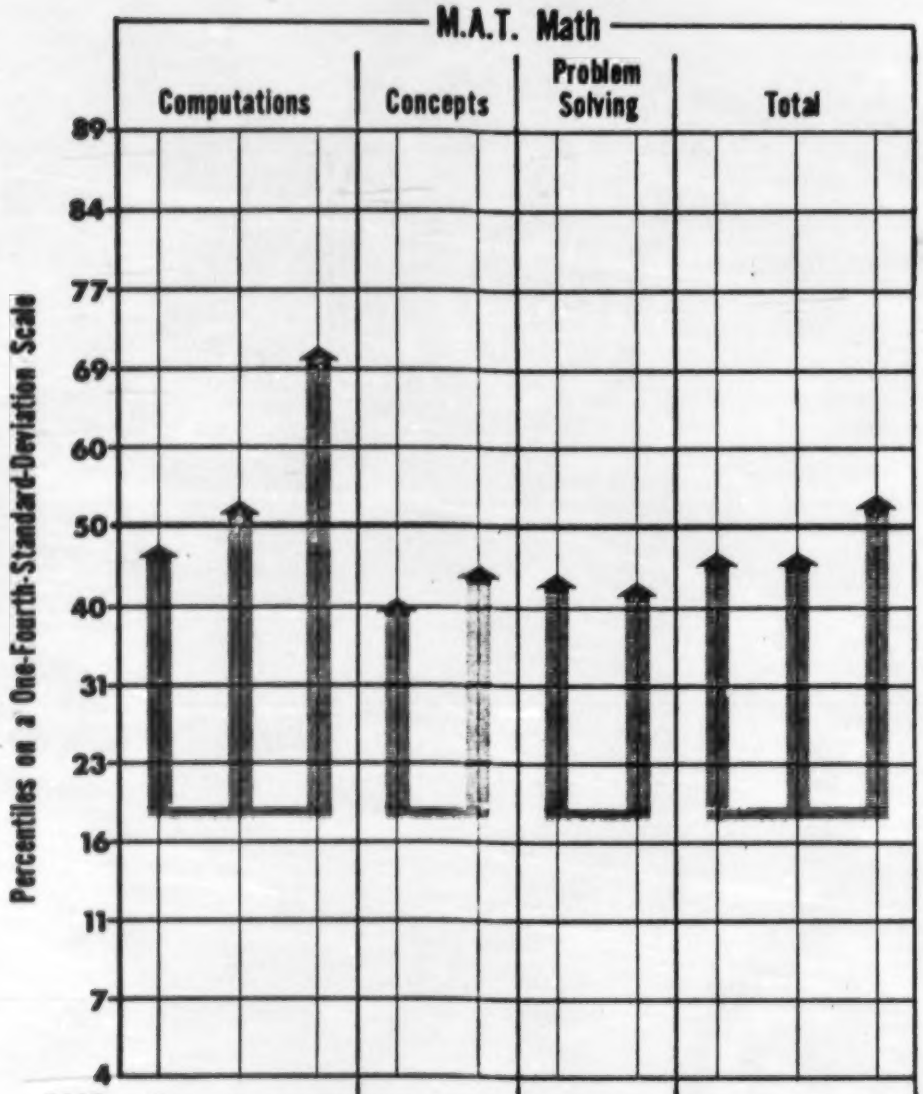


Table 7

Norm-Referenced Gains on MAT Math Subscales by Grade
for EK Low-Income, DI Follow Through Students

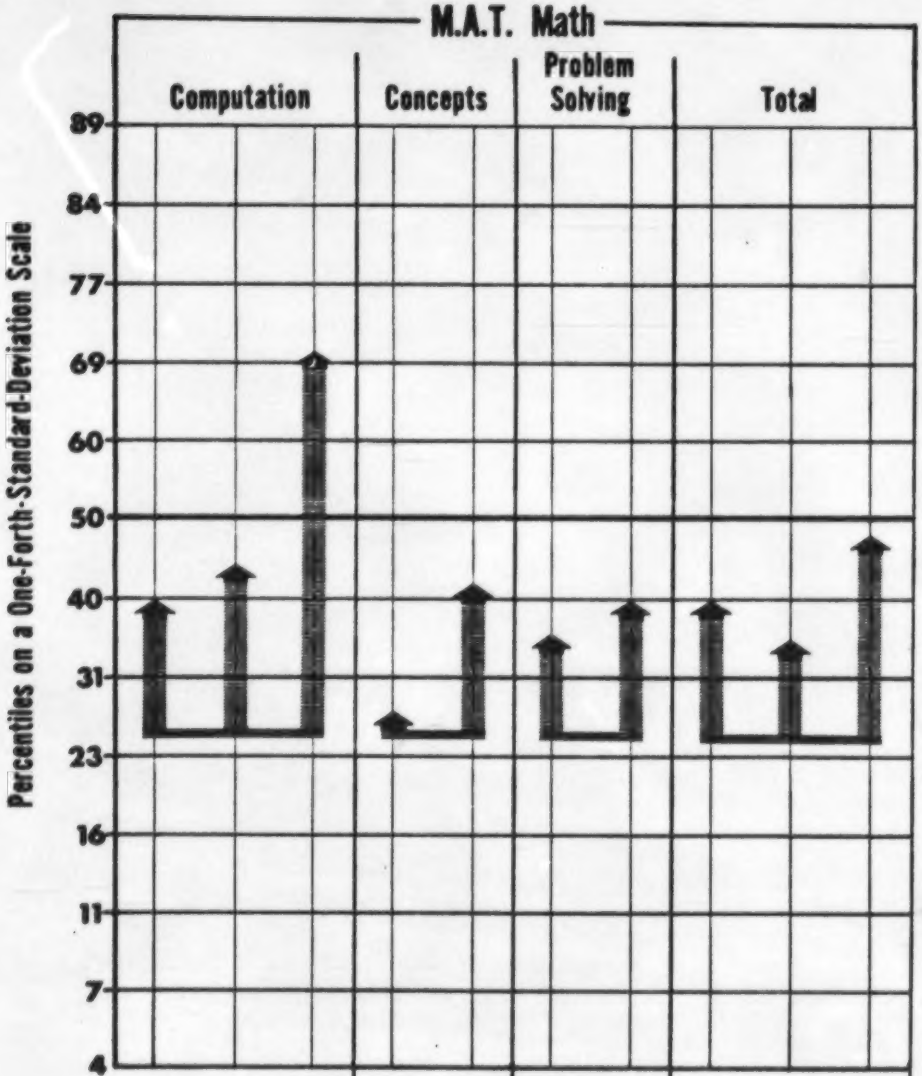


POST	1*	2	3	2	3	2	3	1	2	3
N	2809	2927	2917	2929	2917	2922	2917	2809	2890	2897
%-tile	48	53	71	41	45	44	43	47	47	54
S. S.	40.8	54.1	71.2	53.5	65.6	55.5	65.8	40.8	56.5	71.4
S. D.	11.0	10.2	12.1	10.8	11.8	11.6	12.3	11.0	10.1	11.6
G. E.	1.8	2.9	4.3	2.7	3.7	2.8	3.4	1.8	2.8	4.0

* Combined Computation and Concepts

Table 8

Norm-Referenced Gains on MAT Math by Grade Level
for EF, Low-Income, DI Follow Through Students



POST	1*	2	3	2	3	2	3	1	2	3
N	1267	1683	2399	1686	2393	1684	2395	1267	1677	2385
%-tile	40	44	70	28	42	37	40	40	36	48
S. S.	38.0	52.3	69.9	51.2	64.2	53.3	65.0	38.0	54.4	70.0
S. D.	10.4	9.0	11.8	9.7	10.8	9.9	11.2	10.4	8.6	11.0
G. E.	1.7	2.7	4.3	2.6	3.7	2.7	3.6	1.6	2.5	3.7

* Combined Concepts and Computation

Figure Captions

Figure 1. Percent of Significant Outcomes for Each Follow Through Sponsor on all Math Subtests of the Metropolitan Achievement Test. (Adapted from Stebbins et al., 1977)

Figure 2. End-of-Third Grade Performance for Follow Through (FT) and Comparison (NFT) Children on Total Math for all Entering-Kindergarten (EK) sites. (Adapted from Stebbins et al., 1977)

Figure 1. Percent of Significant Outcomes for Each Follow Through Sponsor on all Math Subtests of the Metropolitan Achievement Test. (Adapted from Stebbins et al., 1977)

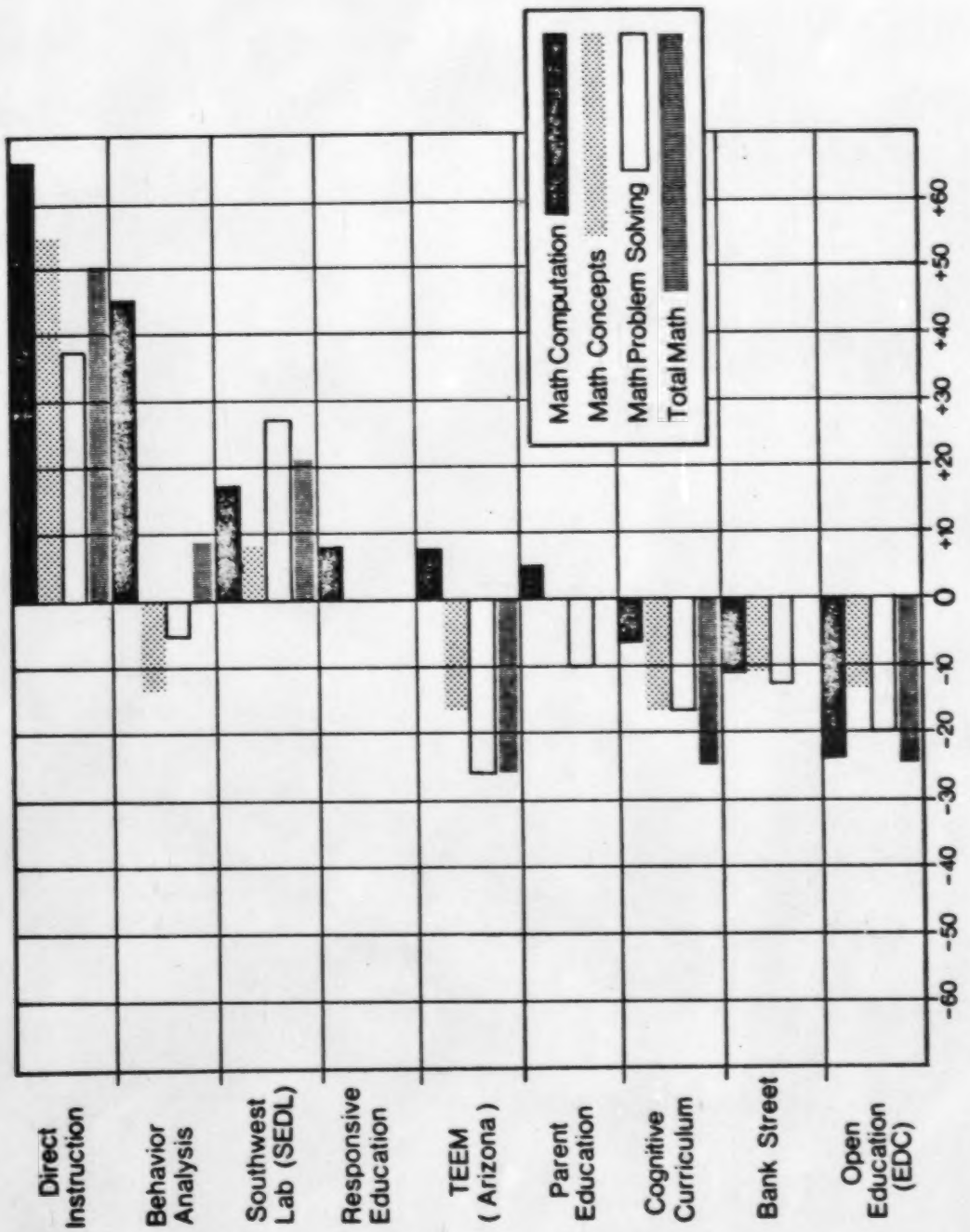
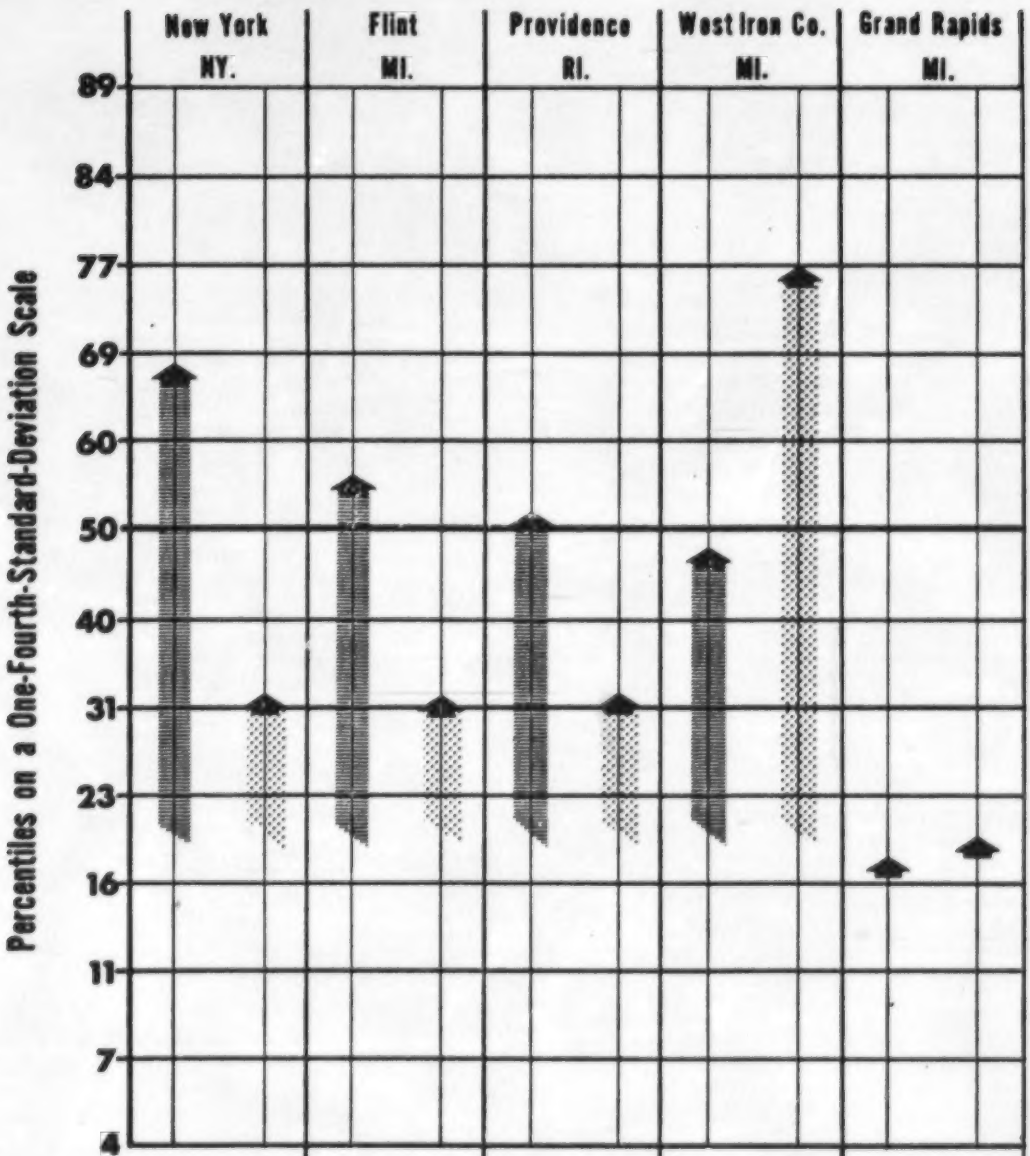


Figure 2

Median Percentile Rankings for Total Math,
Direct Instruction EK sites, cohort III[a]



	New York NY.		Flint MI.		Providence RI.		West Iron Co. MI.		Grand Rapids MI.	
N	FT	Comp	FT	Comp	FT	Comp	FT	Comp	FT*	Comp
%-tile	68th	28th	56th	32nd	52nd	32nd	48th	77th	18th	20th
S. S.	75.8	64.1	72.3	65.0	71	65	70.3	80.2	60.3	61.2

[a] adapted from Stebbins et al., 1977, Vol. IVA, p. 276

* not implementing Direct Instruction model